DOCUMENT RESUME

ED 367 694 TM 021 153

AUTHOR	Murthy, Kavita
TITLE	Commonality Analysis for the Regression Case.
PUB DATE	Apr 94
NOTE	15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 6, 1994).
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Correlation; Heuristics; Higher Education; *Prediction; Predictor Variables; *Regression (Statistics); *Statistical Analysis; Undergraduate Students
IDENTIFIERS	*Commonality Analysis; Statistical Analysis System; *Variance (Statistical)

ABSTRACT

Commonality analysis is a procedure for decomposing the coefficient of determination (R superscript 2) in multiple regression analyses into the percent of variance in the dependent variable associated with each independent variable uniquely, and the proportion of explained variance associated with the common effects of predictors in various combinations. Commonality analysis is an attempt to understand the relative predictive power of regressor variables, both individually and in combination. Despite their utility, these methods have not been used with great frequency, perhaps because these methods are not fully automated in commonly used statistical packages. This paper explores the applications of commonality analyses by reanalyzing data from a study of the relationships among anger and stress in predicting depression among 247 undergraduate students. This data set is employed as a heuristic to make the discussion more accessible. In addition, a Statistical Analysis System (SAS) computer program procedure for obtaining all possible R superscript 2 values is discussed as an efficient method of implementing the required analyses. Two tables and one chart are appended. (Contains 13 references.) (Author/SLD)

Commonality Analysis for the Regression Case

Kavita Murthy

Texas A&M University 77843-4225

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 6, 1994.

ABSTRACT

Commonality analysis is a procedure for decomposing R^2 in multiple regression analyses into the percent of variance in the dependent variable associated with each independent variable uniquely, and the proportion of explained variance associated with the common effects of predictors in various combinations. Commonality analysis is an attempt to understand the relative predictive power of regressor variables, both individually and in combination. Despite their utility, these methods have not been used with great frequency, perhaps because these methods are not fully automated in commonly used statistical packages. This paper explores the applications of commonality analysis by re-analyzing data from a study of the relationships among anger and stress in predicting depression among undergraduate students. This data set is employed as a heuristic to make the discussion more accessible. In addition, a SAS computer program procedure for obtaining all possible R^2 values is discussed as an efficient method of implementing the required analyses.

It has been increasingly recognized that discarding variance by converting intervally-scaled variables into nominally-scaled variables is not good research practice. As Kerlinger (1986, p. 558) explains,

> . . .partitioning a continuous variable into a dichotomy or trichotomy throws information away. . . To reduce a set of values with a relatively wide range to a dichotomy is to reduce its variance and thus its possible correlation with other variables. A good rule of research data analysis, therefore, is Do not reduce continuous variables to partitioned variables (dichotomies, trichotomies, etc.) unless compelled to do so by circumstances or the nature of the data (seriously skewed, bimodal, etc.).

Kerlinger (1986, p. 558) notes that the variance is the "stuff" on which all analysis is based. Discarding variance by categorizing variables amounts to "squandering of information" (Cohen, 1968, p. 441). Pedhauzer (1982, p. 453) agrees that, "Categorization leads to a loss of information, and consequently to a less sensitive analysis."

Similarly, Humphreys and Fleishman (1974, p.468) note that categorizing variables in a non-experimental design using and ANOVA analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong."

In fact, the practice of discarding variance on intervally scaled predictor variables to perform ANOVA, ANCOVA or MANOVA analyses creates problems in most cases. As Cliff (1987, p. 130) notes:

> Think of the persons near the borders. Some who should be highs are actually classified as lows and vice versa. In addition, the "barely highs," are classified the same as the "very highs," even though they

> are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less.

Others (Cohen, 1968, p. 441) have noted that the use of ANOVA-type methods to analyze data reduces the reliability of most variables considered in the design, inflates Type II error probability, discards important information, and distorts the distribution shapes of and relationships among certain variables. These various realizations have lead to less frequent use of OVA methods, and to more frequent use of multiple regression (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1982).

Multiple regression is a statistical technique that offers a method for determining the weights that should be used to obtain the most accurate linear prediction of a criterion from several predictors (Allen & Yen, 1979). Researchers in the social sciences who utilize multiple regression methods in their data analyses typically examine only several aspects of multiple regression results. For example, they usually report the magnitude of a statistically significant multiple regression relationship in terms of the coefficient of determination (R^2), or the extent to which variance in the dependent variable is "explained" by various predictors.

It is somewhat rare for researchers to further decompose the R^2 to determine unique and non unique contributions of the independent variables to prediction of the criterion variable (Siebold & McPhee, 1979). Commonality analysis offers a useful method for partitioning variance because it does not depend upon a priori knowledge of the influence of predictors. Commonality analysis examines all possible orders of entry of the predictors into the model, and the predictors essentially fall where they may. Siebold and McPhee (1979) also argue that :

> Advancement of theory and the useful application of research findings depend not only on establishing that a relationship exists among predictors and the criterion, but also upon determining the extent to

which those independent variables, singly and in all possible combinations, share variance with the dependent variable. Only then can we fully know the relative importance of independent variables with regard to the dependent variable in question. (p.355)

The present paper explains how commonality analysis can be conducted using a particular SAS procedure and some simple computations. To make the discussion concrete, actual data involving undergraduates perceptions of stress and anger as it is related to depression are used for heuristic purposes.

Purpose of Commonality Analysis

The purpose of commonality analysis is to partition a squared multiple correlation into elements associated with each regressor variable and into elements associated with each possible combination of regressors. Commonality analysis generates these elements such that the sum of all elements equals the squared multiple correlation. It is also required that the sum of all elements associated with a single variable is equal to the squared simple correlation of that particular variable with the dependent variable. For the two-predictor case, these relationships can be expressed as:

$$R^2_{y.12} = U_1 + U_2 + C_{12}$$

where R^2_{y12} is the squared multiple correlation of Y with variables one and two, U_1 is the "uniqueness" or unique contribution of variable one to the squared multiple correlation, U_2 is the unique contribution of variable two, and C_{12} is the common element or commonality of variable 1 and 2, or the proportion of variance in Y predictable using either variable 1 or variable 2. As a result, for this case three variance components can be derived from the R^2 of the model, namely U_1, U_2, and C_{12}.

The number of possible unique and commonality components is exponentially determined and can be derived as the difference between the total number of components and the number of unique components, or $(2^p - 1) - P$ where p is the number of independent variables examined. As such, the number of components or variance partitions increases very rapidly as additional predictors are considered.

The rules for calculating the unique and commonality components are straightforward polynomial expressions developed by Mood and Wisler in 1969. However, as the number of independent variables increases, the complexity of the respective component calculations also increases. Table 1 presents the formulas for two, three and four variable models. Seibold and McPhee (1979) offer the formulas for a five variable model.

Insert Table 1 about here

Commonality analysis requires every possible R^2 value for all variable combinations. SAS provides a useful program (PROC RSQUARE) that will print out in ascending order the R^2 values for all possible combinations of the independent variables in the model. This SAS routine makes commonality analysis much simpler, since the calculation of the required R^2 values is fully automated.

The obtained R^2's are then used to determine all unique and common effects, by substituting them into the appropriate commonality formulas listed in Table 1. This can be easily accomplished by using any spreadsheet program to implement the appropriate combinaiton of formulas.

Once the variance components have been determined, the results can be then arranged in a summary table that is easy to interpret and allows for inspection of the total variance associations with each independent variable (column totals) as well as specific unique and common effects (row entries). Another check on the analysis is that the sum of all unique and commonality values should equal the R^2 value of the regression model when all the independent variables are entered into the model.

Heuristic Example

Data from a previous study involving the relationship between anger and stress in predicting depression among undergraduate students can be employed to illustrate the steps in the process of conducting a commonality analysis. The reader may want to refer to this study which was published by Clay, Anderson, and Dixon (1993) in the September/October issue of the Journal of Counseling and Development. Within this study, 247 undergraduates completed three paper and pencil questionnaires assessing stressful life events, depression, and anger expression. The anger expression instrument yielded three different subscales; anger in (IN) anger out (OUT) and anger control (CONT), while the stressful life events instrument yielded only one score (STS).

The results from this study concluded that anger in and stressful life events were significantly related to depression, and that anger out and anger control were not. Thus, the authors decided to eliminate anger out and anger control from further analyses and just focus on the other two variables (STS & IN). But because of the high degree of correlation between all of these predictor variables, (IN, OUT, CONT, &STS), commonality analysis can be used to determine the unique and common components of these variables so that a more accurate explanation in predicting depression can be obtained.

The first step is to obtain the 15 equations necessary for computing the unique and commonality components of a 4-variable model. These equations are obtained from Table 1.

The next step is to then extract all R^2 values from the SAS printout (see appendix A) and substitute these accordingly into the 15 equations. Appendix A presents the R^2 values for all possible combinations of the predictors in this data set.

Third, the appropriate formulas are then applied to the various R^2 values using any calculator or spreadsheet program. For example:

$$U_1\text{ (STS)} = -R^2(234) + R^2(1234)$$

$$= -.15977 + .30414$$

$$= .14437$$

Therefore, the unique contribution of the variable, stress (STS), to the proportion of total dependent variance explained was .14437, or approximately 14%. In addition, the commonality between stress (STS, U_1) and anger in (IN, U_2) can be computed as:

$$C_{12}\text{ (STS/IN)} = -R^2_{(34)} + R^2_{(134)} + R^2_{(234)} - R^2_{(1234)}$$

$$= -.01953 + .20307 + .15977 - .30414$$

$$= .03917$$

Thus, the common variance of the model shared by stress (STS) and anger in (IN) is .03917, or approximately 4%.

The fourth step is to arrange these obtained values into a commonality analysis summary table, like the one presented in Table 2. Once in tabular form, the previously mentioned checks on the data can be performed. For example, summing down columns for each independent variable will yield the R^2 of the regression model in which that independent variable is the only variable entered into the model. Another check is that the sum of all unique and commonality values should equal the R^2 value of the regression model when all the independent variables are entered into the model.

Insert Table 2 about here

Discussion

The commonality summary table presented in Table 2 indicates that the unique predicted variance contribution of the predictor, stress, is approximately 14% (.14437) and its total commonality variance with one or more of the other predictors is approximately 5% (.04923). In this particular example, the variable, stress, is the dominant factor in predicting depression with college students. In addition, anger in, seems to be another helpful predictor of depression because it uniquely contributed approximately 10% (.10107) of the variance and its total commonality variance is approximately 4% (.03583). Consequently, the remaining variables, (anger out and anger control), offer little unique contribution to the variance (.00016, .01067, respectively).

Also, some instances of negative commonalities may occur, as in this particular study with C124, C24, C234, C1234, as reported in Table 2. This result can be "counter-intuitive since the result could be taken to mean that . . . predictor variables have in common the ability to explain less than 0% of the variance" (Thompson, 1985, p. 54). However, the presence of negative commonalities is typically attributable to so-called suppressor effects, which might have been the case for this study (Beaton, 1973).

Commonality analysis is an attempt to understand the relative predictive power of the regressor variables both individually and in combination. This capability offers distinct advantages over more frequently used, traditional, types of analyses such as ANCOVA or stepwise regression. Commonality analysis is straightforward and easy to calculate when no more than four independent variables are involved, with the assistance of the SAS PROC RSQUARE procedure. As such, commonality analysis can be, and should be used more frequently in educational and social science research to partition the variance of the dependent variable into its constituent predicted parts.

References

Allen, M.J. & Yen, W.M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company.

Beaton, A.E. (1973). Commonality. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 111 829)

Cliff, N. (1987). Analyzing multivariate data San Diego: Harcourt Brace Jovanovich.

Cohen, J. (196[illegible]. Multiple regression as a data-analytic system.Psychological Bulletin, 70, 426-443.

Elmore, P.B. & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Goodwin, L.D. & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Humphreys, L.G. & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual differences variables. Journal of Fducational Psychology, 66, 464-472.

Kerlinger, F.N. (1986). Foundations of behavioral research (2nd ed.) New York: Holt, Rinehart and Winston.

Mood, A.M. (1969). Macro-analysis of the American educational system. Operations Research, 17, 770-784.

Rowell, K. (1991, January). Partitioning predicted variance components into constituent parts: How to conduct a commonality analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

Seibold, D.R. & McPhee, R.D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. Human Communication Research, 5, 355-365.

Thompson, B. & Miller, J.H. (1985, February). A multivariate method of commonality analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 263 151)

Wilson, V.L. (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 222 522)

Table 1

Formulas for Unique and Commonality Components of Variance

Two Independent Variables

$U1 = -R^2(2) + R^2(12)$
$U1 = -R^2(1) + R^2(12)$

$C12 = -R^2(1) + R^2(2) - R^2(12)$

Three Independent Variables

$U1 = -R^2(23) + R^2(123)$
$U2 = -R^2(13) + R^2(123)$
$U3 = -R^2(12) + R^2(123)$

$C12 = -R^2(3) + R^2(13) + R^2(23) - R^2(123)$
$C13 = -R^2(2) + R^2(12) - R^2(23) - R^2(123)$
$C23 = -R^2(1) + R^2(12) - R^2(13) - R^2(123)$

$C123 = -R^2(1) + R^2(2) - R^2(3) - R^2(12) - R^2(13) - R^2(23) + R^2(123)$

Four Independent Variables

$U1 = -R^2(234) + R^2(1234)$
$U2 = -R^2(134) + R^2(1234)$
$U3 = -R^2(124) + R^2(1234)$
$U4 = -R^2(123) + R^2(1234)$

$C12 = -R^2(34) + R^2(134) + R^2(234) - R^2(1234)$
$C13 = -R^2(24) + R^2(124) + R^2(234) - R^2(1234)$
$C14 = -R^2(23) + R^2(123) + R^2(234) - R^2(1234)$
$C23 = -R^2(14) + R^2(124) + R^2(134) - R^2(1234)$
$C24 = -R^2(13) + R^2(123) + R^2(134) - R^2(1234)$
$C34 = -R^2(12) + R^2(123) + R^2(124) - R^2(1234)$

$C123 = -R^2(4) + R^2(14) + R^2(24) + R^2(34) - R^2(124) - R^2(134) - R^2(234) + R^2(1234)$
$C124 = -R^2(3) + R^2(13) + R^2(23) + R^2(34) - R^2(123) - R^2(134) - R^2(234) + R^2(1234)$
$C134 = -R^2(2) + R^2(12) + R^2(23) + R^2(24) - R^2(123) - R^2(124) - R^2(234) + R^2(1234)$
$C234 = -R^2(1) + R^2(12) + R^2(13) + R^2(14) - R^2(123) - R^2(124) - R^2(134) + R^2(1234)$
$C1234 = -R^2(1) + R^2(2) + R^2(3) + R^2(4) - R^2(12) - R^2(13) - R^2(14) - R^2(14) - R^2(23) - R^2(24) - R^2(34) + R^2(123) + R^2(124) + R^2(134) + R^2(234) - R^2(1234)$

Table 2

Commonality Analysis Summary Table

Component	1 STRESS	2 ANGER IN	3 ANGER OUT	4 ANGER CONTROL
U1	.14437			
U2		.10107		
U3			.00016	
U4				.01067
C12	.03917	.03917		
C13	.00241		.00241	
C14	.00242			.00242
C23		.00052	.00052	
C24		-.00452		-.00452
C34			.0027	.0027
C123	.00203	.00203	.00203	
C124	-.00113	-.00113		-.00113
C134	.00451		.00451	.00451
C234		-.00006	-.00006	-.00006
C1234	-.00018	-.00018	-.00018	-.00018
TOTAL	.1936	.13690	.01210	.01440
U	.14437	.10107	.00016	.01067
C	.04923	.03583	.01194	.00373

Appendix A

R-Squares of Stress and Anger Expression as Predictors of Depression

Number of Predictores in Model	R-Square	Variables in Model
1	.19360	1 STRESS
	.13690	2 ANGER IN
	.01210	3 ANGER OUT
	.01440	4 ANG CONTROL
2	.29061	1 2
	.20239	1 4
	.19692	1 3
	.15719	2 4
	.14669	2 3
	.01953	3 4
3	.30398	1 2 4
	.29347	1 2 3
	.20307	1 3 4
	.15977	2 3 4
4	.30414	1 2 3 4